

Minimalism and Its Implications for Language Acquisition and Learnability

Norio Suzuki

1. INTRODUCTION

One of the main objectives of generative grammar is to explain how the child is able to acquire her native language in a relatively short time on exposure to relatively degenerate language data. The central concept is Universal Grammar (UG) in conjunction with the notion of the faculty of language (FL). UG is assumed to be a description of FL at the initial state; that is, at the time of birth. FL is one of a number of biological organs human beings are endowed with. Biologically speaking, it can be considered on a par with such familiar components as the visual capacities and walking abilities. In minimalist terms, UG can be said to consist primarily of universal computation (Merge, Agree, and Move (combining Merge and Agree)) and a small number of (morpholexical) parameters that define crosslinguistic variation. (UG was considered to consist of universal principles and parameters in GB terms. But the role of principles has undergone a radical change with the transition from a principled to an un-principled syntax (Epstein 1999) and with the dominance of derivational approaches over representational ones; Move, for example, is no longer free, but must be motivated.) Parameters are restricted to functional domains of the lexicon, so,

lexicons aside, there is only one human language.

Given the logical problem of language acquisition, the language acquisition device (LAD) should be equated with UG, that is, the initial state of FL. This assumption is based on Chomsky's (1975) notion of instantaneous acquisition. The latter is, of course, in conflict with the empirical findings in the literature. (Acquisition cannot be accomplished only with properties of the initial state of FL.) Rather, the actual fact dictates that LAD itself should change with the transition of time, the notion of LAD somehow corresponding to a successive series of states of FL. The main reason that Chomsky does not take this latter position seems to come from the requirement of idealization, which is quite common in the natural sciences. Abandonment of idealization would have made the problem of explanation an insurmountable one. Moreover, the 'logical' properties of the problem can be seen to already invite such a notion as instantaneous acquisition.

As for access to UG, the availability of UG in second language acquisition (L2A) has been among the major topics in the literature. Basically three options have been pointed out: full access, partial access, and no access. Full access approaches to L2 acquisition assume that interlanguage systems are constructed exclusively within the narrow hypothesis space made available by UG. According to the partial access approach, on the other hand, L2 systems arise largely as a result of non-language-specific abilities of the mind, with only some limited access to UG, the latter being presumably restricted to those aspects of UG which can have a direct instantiation in the second language learner's first language (L1). Lastly, no access approaches claim that, in contrast to L1

acquisition, domain-specific knowledge plays no role in L2 acquisition (Dekydtspotter, Sprouse, and Thyre 1999/2000).

2. SOME GENERAL DISCUSSION

On the assumption that one who has acquired her first language (presumably with a certain kind of concomitant physiological hardwiring and neural (re) composition) keeps active universal computation, which constitutes the major part of UG, partly through her linguistic performance (production, processing, etc.; based on the I-language system (provided by FL/UG) turned active upon exposure to experience), it can at least be safely said that ‘parametrized’ (a la her first language) computation should be available to the second language learner. (The notion that those who have acquired their first language ‘keep’ universal computation ‘active’ contrasts with the case of those who have been deprived of any exposure to language altogether in the critical period. In the latter case the development of FL simply does not take place, the component of universal computation and open parameters being left ‘untriggered and, hence, inert’ as they are.) Thus the second language learner can have access to universal computation, which is ‘active’ partly through her linguistic performance and, presumably, for other reasons bearing on problems of physiology and neurology, but which is parametrized according to her native language. So of the three options concerning access to UG (noted in the Introduction), which one is the most plausible? Before getting on to the answer, let us consider precisely what it means to have access to ‘parametrized’ computation.

It has been assumed that certain Scandinavian languages have a grammatical operation called ‘object shift,’ while such languages as English do not. Chomsky (2000) has proposed that the light verb *v* has an EPP-feature in the structures where object shift applies (see Holmberg 1999 for a PF-related view of the phenomenon).¹ Note that the presence of an EPP-feature entails the application of Move (of the object, in this case), including Agree. Although English does not instantiate object shift (for different views see Johnson 1991, Koizumi 1995, and Lasnik and Saito 1991) as language fact, it does not mean that (native) speakers of English do not know Move, Move being part of universal computation and English containing constructions instantiating Move. Only the light verb *v* does not have an EPP-feature in English, a morpholexical fact. Only English happens not to have movement of the object into the outer Spec of *v*P. The fact remains that both Scandinavian and English speakers have exactly the same universal computational system, only the ‘places’ of its applications being different depending on the differences in the morpholexical facts involved in the languages. I presume that the major part of what it means to have access to ‘parametrized’ computation lies in facts of the sort discussed above.

Then, abstracting away from differences in the ‘positions’ of its application, which in turn follow from different morpholexical facts (‘parametrization’), all adult (in the sense of having attained the ‘steady state’) human beings can be said to have the same universal computational system. So we are now in a position to answer the above question. Since we have assumed that UG consists primarily of universal computation and a small number of morpholexical parameters, the latter of which are open at the initial state, the no

access approach is immediately ruled out. We are very close to full access, but there is a large difference in the domain of parameters, ‘open’ vs. ‘fixed’ status corresponding to the initial vs. steady state, respectively. So it might be safe to say that we have partial access to UG in second language acquisition (not committing ourselves for the time being to the problem of ‘non-language-specific abilities of the mind,’ for example (see Dekydtspotter et al. 1999/2000)), but it may be said, on the other hand, that we can almost have full access to it for practical purposes.

Note that, investigating the role of domain-specific knowledge in the syntax-semantics interface of English-French interlanguage grammars within the framework of Chomsky’s (1995) Minimalist Program, Dekydtspotter, Sprouse, and Anderson (1997) observe that “If the Minimalist conception of grammatical knowledge proves correct, the debate on whether domain-specific knowledge in L2 acquisition is derivable from the L1 or directly from UG is essentially moot, because access to the computational principles deriving the L1 is indistinguishable from access to UG” (Dekydtspotter et al. 1997: 299). Their observation may amount to the claim that since it is true that the adult second language learner is given L1 when faced with the problem of acquiring L2, it should be the case that she has full access to UG, given minimalist assumptions. In a nutshell, (the essential component of) UG (crucial to acquisition) is involved in L1’s various aspects (competence and performance), including universal computation in particular. They conclude that “the moment one avails oneself of domain-specific knowledge in constructing an L2 grammar, one faces UG as a function of Minimalist design” (Dekydtspotter et al. 1997: 309). We return later to the problem of

a possible distinction between full and partial access to UG in second language acquisition in more specific and concrete terms.

Then why are there such considerable differences in degree of success between first and second language acquisition: 'perfect' mastery in the former and noticeable individual differences in the latter? Here I assume that FL/UG provides general properties of the lexicon, in addition to universal computation and open morphological parameters (to be fixed on exposure to primary linguistic data). On a par with the cases of universal computation and morphological parameters, I assume that general properties of the lexicon are kept 'active' in those who have acquired L1 (in contrast to the case of language deprivation) through performance and for reasons of physiology and neurology. I assume that a number of differences between L1 and L2 acquisition lie in the acquisition of the lexicon, universal computation being (virtually) the same in both cases. Note that it should be the case that the emergence of physiological hard-wiring and neural (re)composition is restricted to the Critical Period. (Presumably, their emergence never occurs beyond the Critical Period.) It should be rather clear that physiology and neurology (and, perhaps, other factors) are involved in the L1 acquisition of the lexicon, thus accounting for the 'rapid' learning of words, superb retainability of vocabulary items, and so forth. These factors, on the other hand, should not have effects on the L2 acquisition of the lexicon, which is a matter beyond the Critical Period. Rather, the task in the L2 case may depend on some forms of the 'general learning strategy,' thus explaining the 'slowness' of acquisition, vast individual differences in degree of success, and so on.

It should be stressed, however, that, as noted, the L2 acquisition

of the lexicon must also be subject to the constraints from general properties of the lexicon provided by UG. And since parameters are restricted to the functional domains of the lexicon as previously noted, the acquisition of the lexicon must involve that of parameters (L2 acquisition thus inviting such notions as ‘parameter resetting’). But it is not entirely clear whether a putative parameter should belong to the lexicon or to the syntax; in the case of word order, for example, if you do not adopt Kayne’s (1994) theory of antisymmetric syntax, you are somehow forced to commit yourself to the ‘head parameter’ (VO vs. OV), which it must be rather difficult to subsume under the lexicon. Suppose that the notion of ‘morpholexical parameter’ is indeed real, so that there are substantive constraints on the notion of a ‘possible parameter.’ Then the ‘head parameter’ cannot count as a possible parameter, its effects having somehow to be captured in some other way. Note that while the ‘head parameter’ per se as it has been standardly formulated could not possibly be stated in a manner other than syntactic, it would perhaps not be an implausible move to adopt Kayne’s (1994) antisymmetric universal base clausal structure (i.e., SVO) and to assume the existence of an ‘obligatory object raising’ past the verb for languages such as Japanese, the relevant ‘object raising’ presumably being motivated by an EPP-feature on v^* or on some other functional head, depending on the final position of the finite verb in ‘overt’ syntax in such languages. Then the problem in deriving the effects of the ‘head parameter’ in a way compatible with the notion of ‘possible morpholexical parameter’ would be whether or not a specific (functional) category has a specific feature (an EPP-feature, here) as a case of ‘language-level’ parametrization, simply

a matter of the language-particular lexicon (see also the implicational relation between verb raising and object shift, due to Holmberg's Generalization (Holmberg 1999, Chomsky 2001)). Notice further that the notion that there is only one human language, lexicons aside, invites us to assume that parameters, which account for crosslinguistic variation, should be included in the lexicon. Then this conceptual argument supports the reality of the 'morpholexical' properties of parameters. (Recall here that, with regard to language, at a given time of your life you are in some state of FL (Chomsky 1995); presumably, even embryos and fetuses can be said to be in a certain state of FL (see Jenkins 2000 for an epigenetic approach to language).)

Keeping this in mind, let us examine a couple of concrete examples from L2 acquisition.

- (1) a. i. *What do you like [_{t_{what}} books]?
- ii. What books do you like _{t_{what}} books?
- b. i. *How many do you have [_{t_{how many}} books]?
- ii. How many books do you have _{t_{how many}} books?

(Although I generally follow the minimalist copy theory of movement, I indicate the original position of the relevant categories by 'trace' for expository purposes.) (1ai,bi) are examples that I have heard a number of university students (whose first language is Japanese) utter when teaching English (as a second/foreign language) to them in the classroom situation. Note the following examples from some other languages; (2) is a well-known French paradigm, (3a) is an example of Russian from Avrutin (1994), and

(4) is a Dutch example taken from Bennis (1986):

- (2) a. Combien de livres a-t-il consultés $t_{\text{combien de livres}}$?
b. Combien a-t-il consultés [$t_{\text{combien de livres}}$]?
‘How many books did he consult?’
- (3) a. Č’ju on vygulival [$t_{\text{č’ju sobaku}}$]?
whose he walk-PAST dog
‘Whose dog did he walk?’
b. *Whose did he walk [$t_{\text{whose dog}}$]?
- (4) Wat heeft hij [$t_{\text{wat voor romans}}$] geschreven?
what has he for novels written
‘What novels has he written?’

Abstracting away from a variety of analyses of these examples largely in GB terms in the literature, the least that can be said from the data in (2–4) above (excluding the English case (3b)) is that it is possible to extract a constituent from a ‘specifier-like’ position in human languages. Then the Japanese-English interlanguage grammars that generate such examples as in (1a,bi), where instances of extraction of a ‘specifier-like’ element can be seen, can be said to be among possible human languages, arguably ‘UG-consistent.’ (As the Japanese translations of (1a,b) ‘Donna hon-ga suki-desu-ka?/Hon-o nansatsu (Nansatsu-no hon-o) motte-imasu-ka?’ may indicate, the reason for the separation of *what* from *books* and *how many* from *books* in (1a,bi), respectively, cannot be sought in the (surface) strings of the relevant sentences in their first language, Japanese.)

Now, by looking into the relevant differences between the English

case (e.g., ‘What books do you like?’ (1aii)) and the Japanese-English interlanguage case (‘*What do you like books?’ (1ai)), the analysis of the latter of which could arguably be extended to the cases from other languages as in (2, 3a, 4) with necessary (presumably, parametric) qualifications, let us examine whether the differences in question can be stated in terms of ‘morpholexical’ parameters that are subsumed under the lexicon, thus maintaining that there is only one human language, lexicons aside. At some stage of the derivation of both of (1ai, ii) we get the following (only the relevant portions given):

(5) [TP you T [_v*P t_{you} like what books]]

Next, when C is merged with TP, it is somehow endowed (at least) with *uQ* (an uninterpretable Q-feature; Chomsky 2000) and *EPP* (an EPP-feature, which induces copying of material from the goal into the local environment of the probe and which is somehow associated with a relevant uninterpretable feature; e.g., with *uQ*, *v**’s uninterpretable ϕ -features, etc. (Chomsky 2000, Pesetsky and Torrego 2001)). Note that the *wh*-phrase has *uWh* (an uninterpretable Wh-feature) and *Q* (an interpretable Q-feature). Then we get the following structure with merged C, both for (1ai, ii):

(6) [C, *uQ*, *EPP*] [TP you T [_v*P t_{you} like what books]]

Under matching of Q-features, Agree holds between the probe C (in virtue of *uQ* in it) and the goal *what* (*books*), which is made active

thanks to the presence of uWh in it, eliminating uninterpretable features (uQ , uWh) that activate them. Since uWh is not matched but deletes under matching of Q -features, I adopt the weaker version of the condition that it *may* delete under Agree (see successive cyclic A-bar movement cases).

This condition could be extended to the case of uninterpretable Case under matching of ϕ -features. (Unwanted cases would be dealt with by FI at the interface or by Boeckx's (2001) assumption that an element with an uninterpretable feature cannot be freed for interpretation and thus remains uninterpretable.) Although it may straightforwardly be pointed out that it should be uWh that makes *what* (*books*) active, it may also be the uninterpretable structural Case of *what books*, which is arguably deleted under agreement with v^* (actually, only 'marked' for deletion and thus present until the stage of CP, given Chomsky's (2001, his (10)) principle to the effect that Phase_1 is interpreted/evaluated at Phase_2 , where Phase_1 is strong and Phase_2 is the next highest strong phase (where a strong phase is CP or v^*P), and adopting Boeckx's (2001: 518) assumption that "as long as an NP has an unchecked Case feature, its feature set is uninterpretable. Once Case is checked, the element is freed for interpretation." Note that the relevant *wh*-element (*what*, *what books*) is arguably in the outer Spec of v^*P (its edge) at the point when the strong phase CP is reached, due to the Phase-Impenetrability Condition (PIC; Chomsky 2000, 2001). (Notice also that v^* may be assigned an EPP-feature only if that has an effect on outcome, even in non-object shift languages (Chomsky 2001: 35). For the notion of 'look-ahead' see the discussion in Stepanov 2001, note 11.)

At the stage of the phase v^*P , the EPP-feature in v^* allows *what* and *what books* to merge at its outer Spec in (1ai, ii), respectively. Raising through the outer Spec of v^*P , the *wh*-phrases land finally in the Spec of CP, so what we will get in the end will be the two-membered \bar{A} -chain ([Spec,C], [Spec, v^*]) and the two-membered A-chain ([Spec, v^*], *what (books)*). Now, what regulates generalized pied-piping that determines the size of the material from the goal to be moved (*what* vs. *what books* in (1ai, ii), respectively)? Trying to account for differences between cases of Condition A of the binding theory and ones of its Conditions B and C with respect to reconstruction, Chomsky (1995:209) proposes a preference principle for reconstruction, such as the following: “Do it when you can (i.e., try to minimize the restriction in the operator position).” Building on this suggestion of Chomsky’s (1995), I tentatively propose an ‘economy’ condition, as in the following:

- (7) Minimize the effects of the EPP-feature.

In terms of markedness based on (7), Japanese, where only checking of features (Agree), but not pied-piping, is involved, is ‘unmarked’ (see Watanabe 1992), while English, where *what books* moves, is the most ‘marked’ case, Japanese-English interlanguage grammars, where only *what* moves, coming in-between. The effects of the EPP-feature are null in the Japanese case, whereas the size of the pied-piped material is variable in the other cases with an EPP-feature, depending on the grammars involved. Notice that the claim that (7) is an economy condition subsumes under it the assumption that UG principles are applied wherever possible, with language-

particular rules used only to ‘save’ certain structures, the former thus being ‘less costly’ than the latter (Chomsky 1995).

Although EPP itself is a feature given by UG (‘an uninterpretable selectional feature’ (Chomsky 2001: 10)), that is, a feature provided by its general properties of the lexicon (so I assume), its ‘distribution’ in structures is a matter of parametric variation, that is, a language-particular matter. The latter fact suggests that ‘effort’ is required to place it in structures; ‘parametric effort,’ in particular, is needed to put it in appropriate places. (The factual cases, though, should all be only ones of ‘parametric effort,’ subsuming ‘word learning’ as well and, perhaps, all other cases of lexical acquisition, on the stronger assumption that all that is ‘innate’ (universal computation, etc.) is already given.) So, from the viewpoint of economy, the mere existence of an EPP-feature in a structure of a language is ‘costly’ since it must already have been put in the particular place of the structure through a specific instantiation of parametric choice, a matter of a language-particular lexicon, but not one of universal computation. Thus we gain some conceptual support for an argument for the postulation of the economy condition (7), ranking the Japanese case as the highest in the EPP economy hierarchy. I assume that EPP is a species of morpholexical feature (drawn from UG’s general properties of the lexicon) and that EPP assignment is a matter of morpholexical parameter-setting, which invites us to take both EPP and its assignment to be problems that should be addressed in terms of a language-particular lexicon (buttressed by UG-given general properties of the lexicon). Hence, the presence vs. absence of an EPP-feature in a given structure is a lexical parametric matter, that is, a problem of the

lexicon.

Then, the relevant difference between Japanese, on the one hand, and Japanese-English interlanguage grammars and English, on the other, can be reduced to the absence vs. presence of an EPP-feature in v^* and C, a strictly lexical matter (i.e., a matter of whether or not a specific category has a specific feature). And the further difference between Japanese-English interlanguage grammars and English (movement of *what* vs. *what books*, respectively) may be sought in an economy (sub-)condition for EPP cases (presumably, derivable from (7)): “Reduce the amount of pied-piping.” The latter might be implemented through economy calculation in terms of the number of ‘words’ pied-piped; the larger the number, the more costly the operation (although the final solution will be seen to make use of the notions of ‘term’ and ‘label’). Then this should not be a syntactic (computational) matter; rather, it must be a problem of the lexicon (so I assume). Throughout universal computation remains the same across all the languages and grammars discussed thus far. (If something along the lines of (7) proves to be correct, it might have some theoretical implications for such problems pertaining to the so-called ‘null subject’ phenomenon as the original problem of accounting for such languages as Italian, the notion of ‘morphological uniformity’ (Jaeggli and Safir 1989), the case of subject-less sentences in child English (Hyams 1986), and the notion of ‘verbal agreement’ in structures without a ‘subject’ (Alexiadou and Anagnostopoulou 1998), among others, their specific analyses aside.)

3. LEARNABILITY

Why do these university students (who produce (1ai, bi), etc.) pass through the state of their FL observed (and go on finally to the target language, English)? What does it mean to have a markedness hierarchy such that Japanese \rightarrow Japanese-English interlanguage grammars \rightarrow English ('unmarked' \rightarrow 'marked') in terms of the economy condition (7) (including the sub-condition concerning the 'amount of pied-piping')? Moreover, it should be somewhat surprising that the students somehow do appear to follow the hierarchy in their L2 acquisition. But we will see shortly that some learnability considerations point to the reality of the condition (7) and its consequences.

Notice that while there are a number of natural languages which exhibit structures similar to (1ai, bi), the students in question can be taken to never have encountered relevant structures in such languages in a meaningful way. They arguably have only been exposed to Japanese, their first language, and (fragments of) English in classroom and, perhaps, other situations. The latter language only instantiates (1aii, bii), etc., but not (1ai, bi). Why has it been possible for them to produce (1ai, bi), etc., in spite of lack of (positive) evidence? (Note that nobody has implicitly or explicitly taught them to produce (1ai, bi), etc.) This may point to the existence of the much-discussed 'innateness' component pertaining to language and invite the confirmation of its existence from the perspective of 'poverty of the stimulus' (i.e., in terms of the property of 'underdetermination' on the part of the input, hopefully, in domain-specific ways), to which I return below.

Let us first view the problem from the learnability point of view. A certain set of learnability hypotheses involving the notion of ‘subset’ relation between the languages considered is known to have been instrumental in describing the binding-theoretic situation between the world’s many languages and (partially) explaining their acquisitional situation (Manzini and Wexler 1987, Wexler and Manzini 1987, Atkinson 2001); the discussion (specifically, the notion of subset relation) being focused primarily on L1 acquisition, but reasonably able to be generalized to the case of L2 acquisition, where (virtually) full access to UG obtains, as noted). Originally, largely in terms of the subset relations between the relevant ‘E-languages’ (sets of sentences) in the sense of Chomsky (1995) (Atkinson 2001), the acquisitional direction has been shown to be from a ‘subset’ language (‘unmarked’) to a ‘superset’ one (‘marked’) (due to the Subset Principle, which requires the learning function to map the input data to that value of a parameter which generates a language that is compatible with the input data and *smallest* among the languages compatible with the input data (Wexler and Manzini 1987, Atkinson 2001)). Later a variety of attempts have been made to remedy the hypothetical acquisitional situation, where ‘computational requirement’ would be implausible, and to make the theory more or less consistent with Chomsky’s (1995) reasonable abandonment of E-language as a legitimate object of scientific inquiry and concomitant concern for I-language (grammar) (inviting the distinction between ‘extensional’ and ‘intensional’ approaches to the problem; see, for example, Wexler’s 1993 attempt to reinterpret his own (with Manzini) original proposal as an ‘intensional’ principle; see also Safir 1987 and Kapur, Lust, Harbert,

and Martohardjono 1993).

While efforts have been made to reinterpret the original formulation of the Subset Principle from the viewpoint of I-language, the Binding Theory itself has seen significant developments in the meantime (e.g., Reinhart and Reuland 1991, 1993; Reuland 2001). Reviewing the (possible) changes that the Subset Principle has undergone thus far and developments of the Binding Theory (the latter theory having been practically the sole rationale for the former principle), Atkinson (2001) somehow brings his discussion to a pessimistic conclusion (although acknowledging the importance of the recent past of the principle), noting that “it seems that no version of the Subset Principle has an application to the core derivations of the Minimalist Program” (p.48). In spite of this overview of the present situation surrounding the Subset Principle, I would like to argue that there does exist at least some residue of the principle on the evidence of some data from second language acquisition, thus pointing to a certain reality of the property of ‘subset’ relation in the acquisitional situation.

In order to continue with my argument for the reality of the notion of subset relation in the acquisitional situation, let me introduce the notion of ‘cue’ with the aid of an example from Lightfoot (1999) within the framework of the cue-based approach to language acquisition. (Note that the latter approach based on cues has been proposed primarily to account for L1 acquisition, but I presume that it may well be applied to the case of L2 acquisition, and also that my analysis involving the notion of subset relation may well go through on approaches other than the cue-based one.) Cues are abstract structures and elements of I-language. The child scans the

linguistic environment for designated structures or cues only in simple syntactic domains, the latter being ‘degree-0 learnability’ of Lightfoot (1991, 1994), relevant primarily to L1 acquisition. The learner seeks cues and may or may not find them, regardless of what the emerging grammar can generate, the success of which is in no way based on the set of sentences that it generates (unlike in input-matching models). And the child’s triggering experience is best viewed as a set of abstract structures manifested in the mental representations which result from parsing sentences, that is, cues, which mainly constitute partial parses, typically lacking some of the information in adult parses (see Lightfoot 1999: 149).

Lightfoot (1999: 153) gives the following example of cue with respect to the verb-second (V2) property of most Germanic languages, in particular:

- (8) $_{\text{SpecC}}[\text{XP}]$ (V2 cue)

The cue in (8) is an abstract structure and an element of I-language. The standard analysis of V2 has it that the finite verb moves to C and an XP that is arbitrary with respect to its grammatical function moves to the [Spec, C] position, the finite verb in the second position playing a crucial role in computing the XP to be in [Spec, C] in the acquisitional situation (see Chomsky 1995 for a suggestion that V2 is a PF phenomenon). The cue (8) must be attested robustly in the primary linguistic data for the child to set the V2 parameter appropriately (Lightfoot 1999, Atkinson 2001).

Let us recapitulate the main point of the discussion. The basic paradigm that has to be explained is the following:

- (9) a. Do you like what books? (Japanese; English version of
‘Donna hon-ga suki-desu-ka?’; ungrammatical as English)
b. *What do you like [_{t_{what}} books]? (Japanese-English
interlanguage grammars; (1ai))
c. What books do you like _{t_{what}} books? (English; (1aii))

In terms of the ‘amount of pied-piping’ in connection with the *wh*-element in (9a–c), there seems to be no way of capturing the distinction between them other than parametrization. So let me propose the following ‘subset’ parameter associated with the economy condition (7):

- (10) Reduce the effects of the EPP-feature to the specified amount of pied-piping, where the specified amount of pied-piping is:
a. 0, or
b. a term, or
c. a label

The option (10a) can presumably be equated with the situation where we have no application of EPP, but I include it among the options associated with the ‘amount of pied-piping’ parameter. (I largely express the situation in (9a) as a case of the non-existence of an EPP-feature, though, while taking the situation to be one where the ‘effects of the EPP-feature’ have been reduced to zero since the ‘amount of pied-piping’ is null.)

A word on some terminology (‘term’ and ‘label’) may be in order. I assume the definition of ‘term’ (Chomsky 1995: 247) and that of

‘syntactic object,’ into which the notion of ‘label’ enters (Chomsky 1995: 243). First, let me point out that ‘a term’ in (10b) is such that it refers to an element (the goal) that has a relevant uninterpretable feature and a relevant interpretable feature, the latter of which corresponds to a relevant uninterpretable feature in the probe that is associated with the relevant EPP-feature appearing in the definition (10), and that ‘a label’ in (10c) should be one associated with ‘a term’ in (10b), the intended meaning being that it is not the label itself (since it is just a ‘label’ without ‘substance’ in the absence of (an) associated term(s)), but the term(s) associated with the label (along with the label in (10c), though) that the operation of generalized pied-piping moves.

Then, for the sake of discussion on some relevant points, let me compare the following two cases:

- (11) a. What do you like t_{what} ?
 b. *What do you like [t_{what} books]? (1ai/9b)

Anticipating the cue-based approach to language acquisition (Lightfoot 1999), let us tentatively take the relevant cue in connection with (10) to be the following both for (11a, b):

- (12) $\text{SpecC}[\textit{what}]$

It is immediately clear that (12) cannot distinguish between (11a, b). Specifically, on the assumption that it is only cues that enter into parameter-setting, the grammatical case in (11a) would lead the learner to expect that (11b) should also be good, since it is also

based on the same cue that has led to a grammatical result and, moreover, is the only input available to the learner in this case. But note that the ‘amount of pied-piping’ parameter in (10) makes reference to notions such as ‘term’ and ‘label’ in its definition. Assuming that the representation of ‘term’ in set-theoretic terms (as is usually the case with the Bare Phrase Structure theory) enters into computational derivations, specifically into the structures of cues, I propose that the relevant cues for (11a, b) should be like the following:

- (13) a. SpecC[*{what <label>, {what <term>}}*<term>] (for (11a))
- b. SpecC[*what <term>*] (for (11b))
- (see *what books: {what <label>, {what <term>, books <term>}}*<term>)

In (13) and elsewhere, I tentatively use a ‘mixed’ notation involving the labeled bracketing notation and the set-theoretic one for ease of exposition to indicate relevant structures and representations. I also insert the status names ‘term’ and ‘label’ for lexical items, etc. in (13) for expository purposes, but minimalist assumptions would exclude them, the status for a specific lexical item, for example, in a structure presumably being able to be determined by general principles and assumptions. I will omit such names in my further exposition, noting that labels do not Merge or Move by themselves (though they may enter into an Agree relation in specific circumstances, presumably as the ‘windows’ for the terms they are associated with). Note also that in (13a) the term ‘component’ consists

of a one-membered set (i.e., *what* <term>; see Stepanov 2001:113 for an instance of such a one-membered term set with a label; see also Chomsky 1995:249, for example, for the phrase structure status of the relevant elements in chains in the Bare Phrase Structure framework).

It can be seen that the cue representations in (13) (to whose validity I turn shortly), which are fragments of I-language, can make a necessary distinction between (11a, b). To make the explanatory foundation more solid, let us suppose that at least the following assumptions (hopefully, following from general principles) should be included among those that regulate phenomena involving the notions of ‘term’ and ‘label’:

- (14) a. Categories that do not project any further have a label.
- b. A chain does not have distinct labels, where distinct labels have a different set of (a) term(s) associated with them.
- c. Movement cannot strand a label with no term(s) associated with it.

(14c) should be so since a label is just a name attached to a set of (a) term(s) and thus without any substance in the absence of (an) associated term(s) (see above). Thus, in a term {x, {x, y}}, where x and y are terminal elements, for example, y must have a label, due to (14a), which also accounts for why the representation of *what* in (13a) must include a label in it (see also (11a)). Assuming the copy theory of movement, and also that {*what* (copy), {*what* (copy)}} is left behind in the original position in (11a), the presence

of a label in (13a) is accounted for by (14c). The absence of a label in the cue representation in (13b) can be accounted for by (14b); otherwise, the relevant chain would have had both $\{what, \{what\}\}$ and $\{what, \{what \text{ (copy)}, books\}\}$, involving two distinct labels. I await future research as to the problem of whether the element in [Spec, C] in (13b) (see (11b)) should count as a ‘category’ in the sense of (14a) (see Chomsky 1995 for discussion on the phrase structure status of categories in chains).

Now we are in a position to begin to deal with the paradigm in (9). But, before going into the specific details of the analysis of relevant points in the paradigm, some general comments on the notions ‘term’ and ‘label’ and a certain operation mediating between a term and its associated label (i.e., Chomsky’s 1973 proposal concerning features; called ‘percolation’ henceforth, for lack of a better term) may be in order. First, terms can safely be said to be able to enter into any kind of general operations with or without their associated label (including Merge, Agree, and Move), but, as noted, labels do not Merge or Move by themselves (though they can do so along with the term(s) they are associated with), and there may be specific circumstances where labels themselves can enter into an Agree relation as the ‘windows’ into the terms they are associated with, the situation with labels definitely being due to its status only as a ‘name’ for its associated term(s). (See Chomsky 1995: 244 for a suggestion that the notion of ‘label’ be dispensed with on the assumption of its unique determination, but I continue to indicate labels throughout for clarity of exposition, as Chomsky 1995 does, noting that my analysis may well go through even without such a notion.)

I claim that this situation arises from the fact that labels are basically without features (perhaps, with the exception of categorial features), so for them to enter into an operation (Agree, as well as Merge and Move, along with the term(s) in the latter two cases) an extra operation (see Chomsky's 1973 proposal concerning features/ 'percolation'), whose application may be parametrized, must apply to let them have relevant (formal, particularly for the purposes here) features that are somehow 'copied' into them from their associated term (basically, a head); thus, the feature(s) necessary for labels to enter into a computational operation should be supplied to them from the head term by 'percolation,' presumably at the time of the application of the relevant operation (Merge, Agree, etc.).² Though this is an extra operation and may always be a reason for parametrization, there may be certain cases where this operation seems (apparently) universally necessary; e.g., the case of y of $\{x, \{x, y\}\}$, where x and y are terminal elements. (Note that this is not surprising in any way since 'percolation' is a 'universal' operation (see the discussion in note 2) and may enter into (apparently) universal processes as well as language-particular ones, the difference between the two cases residing only in 'statistics' (see the discussion on the notion of 'lexical parameter' in Suzuki 1997, chapters 3 and 4).) Notice that 'percolation' must have applied in (11a) since the label (*what*) has entered into an Agree relation (and so forth), while it does not in (11b) (see (13a, b)), indicating that the former must be more 'marked' than the latter due to the extra step of feature percolation. I would like to await future support from experimental research, for example, for the markedness prediction of (11a, b). And anticipating part of the analysis of (9) somewhat,

the Japanese case (9a) should be one where the term *what* (*donna*) is involved in an Agree application, but not the label *what* (*donna*) with the terms *what* (*donna*) and *books* (*hon*) associated with it, since the latter situation would have needed an extra operation of ‘percolation’ (assuming the Japanese case not to involve Move or the effects of EPP (see (10a))).

By way of accounting for the paradigm in (9) (see also (1ai, ii)), let us tentatively look at a hypothetical, though plausible, acquisitional situation concerning the cases discussed thus far. As noted above, the present discussion is couched in the framework of the cue-based approach to language acquisition (Lightfoot 1999), which has primarily been proposed to account for L1 acquisition cases, but which I believe may be extended to the case of L2 acquisition. The crucial input for acquisition, specifically for parameter-setting, on the cue-based approach is ‘cues,’ that is, fragments of I-language, under the assumption that there should be a certain amount of ‘preanalysis’ ability (presumably, due to UG) on the part of the learner (Atkinson 2001, Wexler and Culicover 1980). Now the relevant cues for Japanese (9a) and English (9c/1aii) may be like the following:

- (15) a. $\text{SpecC}[\text{e}]$ (for (9a))
 b. $\text{SpecC}[\{\textit{what}, \{\textit{what}, \textit{books}\}\}]$ (for (9c/1aii))

‘e’ in [Spec, C] of (15a) indicates that no terminal elements have moved into [Spec, C]. On the cue-based approach to language acquisition/parameter-setting, the child is assumed to manipulate only ‘fragments’ of I-language/grammar, but not whole structures/

sentences in their L1 acquisition. The relevant fragments of I-language are (15a, b) for Japanese and English, respectively (both in the cases of L1 and L2 acquisition, I assume). The usual situation may be that the learner fixes the relevant parameter at the relevant value, presumably only after enough exposure to relevant linguistic data; that is, ‘robust’ evidence is needed for parameter-setting (Lightfoot 1999).

Now, the learner of Japanese may be exposed to fragments of I-language of the type (15a) (contained in such (simple) sentences as ‘Donna hon-ga suki-desu-ka? (What books do you like?)’) an enough number of times to set the ‘amount of pied-piping’ parameter (see (10)) at the value a , while the English learner attains enough exposure to fragments of I-language of the type (15b) (contained in such ordinary sentences as ‘What books do you like?’) to fix the parameter at the value c . Let me elaborate on the latter situation, where the operation of ‘percolation’ should be involved (see above; Chomsky 1973). (First, note that the simple presence of an EPP-feature associated with uQ in the probe C would identify a term with uWh and Q as a goal (e.g., *what* in (6)), which stage the (L2) learner (of English) would pass through even without positive evidence of the strict type, presumably due to the subset parameter in (10).) Now, the (marked) operation of ‘percolation’ (provided by FL/UG; see note 2) somehow becomes activated in the child robustly exposed to such data as ‘What books do you like?’ and so forth, copying into the label (the first *what* in (15b)) the relevant features (including at least uWh and Q) from the term/head (the second *what* in (15b)), enabling the label (now a goal) to enter into an Agree relation with the probe C , and at the same time or already

setting the ‘amount of pied-piping’ parameter at the value *c*. (Or rather, the ‘content’ of parameter-setting in this case may reside in the application of ‘percolation’ and the concomitant assignment of the status ‘goal’ to the label for the purposes of Agree here.)

4. CONCLUSION

I have given in this paper ‘Part 1’ of my discussion on L2 acquisition and learnability, which I must admit should necessarily be incomplete in its exposition (I can safely say, though, that I have already presented in the paper (almost) all the ingredients needed for my explanation of the data here), in the framework of the cue-based approach to language acquisition (Lightfoot 1999) on the basis of a number of minimalist assumptions (Chomsky 1995, 2000, 2001). I hope to be able to present a sequel to this paper in not so distant a future.

[To be continued]

NOTES

1. The Extended Projection Principle (EPP) was originally proposed by Chomsky (1981) to account for the presence of a ‘subject’ within a clausal structure. This has recently been generalized to such categories as *v* and C in addition to T.
2. Categorical features alone might be sufficient for Merge to apply, at least for the purposes here. I would assume that ‘percolation’ (though only a descriptive term) is a ‘universal’ operation (hopefully, derivable from Merge or Agree in some way or other) provided by FL/UG (see Suzuki 1997, chapters 3 and 4, where a universal operation enters into a language-particular process which may lead to

a ‘lexical’ parameter). Like Move, ‘percolation’ is ‘costly’ and a ‘marked’ process.

REFERENCES

- Alexiadou, A. and E. Anagnostopoulou (1998) “Parametrizing AGR: Word Order, V-Movement and EPP-Checking,” *Natural Language & Linguistic Theory* 16, 491-539.
- Atkinson, M. (2001) “Learnability and the Acquisition of Syntax,” in S. Bertolo, ed., *Language Acquisition and Learnability*, Cambridge University Press, New York.
- Avrutin, S. (1994) “The Structural Position of Bound Variables in Russian,” *Linguistic Inquiry* 25, 709-727.
- Bennis, H. (1986) *Gaps and Dummies*, Foris, Dordrecht.
- Boeckx, C. (2001) “Scope Reconstruction and A-movement,” *Natural Language & Linguistic Theory* 19, 503-548.
- Chomsky, N. (1973) “Conditions on Transformations,” in S. Anderson and P. Kiparsky, eds., *A Festschrift for Morris Halle*, Holt, Rinehart & Winston, New York.
- Chomsky, N. (1975) *Reflections on Language*, Pantheon Press, New York.
- Chomsky, N. (1981) *Lectures on Government and Binding*, Foris, Dordrecht.
- Chomsky, N. (1995) *The Minimalist Program*, MIT Press, Cambridge, Massachusetts.
- Chomsky, N. (2000) “Minimalist Inquiries: The Framework,” in R. Martin, D. Michaels, and J. Uriagereka, eds., *Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik*, MIT Press, Cambridge, Massachusetts.
- Chomsky, N. (2001) “Derivation by Phase,” in M. Kenstowicz, ed., *Ken Hale: A Life in Language*, MIT Press, Cambridge, Massachusetts.
- Dekydtspotter, L., R. A. Sprouse, and B. Anderson (1997) “The Interpretive Interface in L2 Acquisition: The Process-Result Distinction in English-French Interlanguage Grammars,” *Language Acquisition* 6, 297-332.

- Dekydtspotter, L., R. A. Sprouse, and R. Thyre (1999/2000) "The Interpretation of Quantification at a Distance in English-French Interlanguage: Domain Specificity and Second-Language Acquisition," *Language Acquisition* 8, 265-320.
- Epstein, S. D. (1999) "Un-Principled Syntax: The Derivation of Syntactic Relations," in S. D. Epstein and N. Hornstein, eds., *Working Minimalism*, MIT Press, Cambridge, Massachusetts.
- Holmberg, A. (1999) "Remarks on Holmberg's Generalization," *Studia Linguistica* 53, 1-39.
- Hyams, N. (1986) *Language Acquisition and the Theory of Parameters*, Reidel, Dordrecht.
- Jaeggli, O. and K. J. Safir (1989) "The Null Subject Parameter and Parametric Theory," in O. Jaeggli and K. J. Safir, eds., *The Null Subject Parameter*, Kluwer Academic Publishers, Dordrecht.
- Jenkins, L. (2000) *Biolinguistics: Exploring the Biology of Language*, Cambridge University Press, New York.
- Johnson, K. (1991) "Object Positions," *Natural Language & Linguistic Theory* 9, 577-636.
- Kapur, S., B. Lust, W. Harbert, and G. Martohardjono (1993) "Universal Grammar and Learnability Theory: The Case of Binding Domains and the 'Subset Principle,'" in E. Reuland and W. Abraham, eds., *Knowledge and Language Volume 1: From Orwell's Problem to Plato's Problem*, Kluwer Academic Publishers, Dordrecht.
- Kayne, R. S. (1994) *The Antisymmetry of Syntax*, MIT Press, Cambridge, Massachusetts.
- Koizumi, M. (1995) *Phrase Structure in Minimalist Syntax*, Doctoral dissertation, MIT, Cambridge, Massachusetts.
- Lasnik, H. and M. Saito (1991) "On the Subject of Infinitives," *CLS* 27, 324-343.
- Lightfoot, D. (1991) *How to Set Parameters*, MIT Press, Cambridge, Massachusetts.
- Lightfoot, D. (1994) "Degree-0 Learnability," in B. Lust, G. Hermon, and J. Kornfilt, eds., *Syntactic Theory and First Language Acquisition: Cross-Linguistic Perspectives Volume 2: Binding, Dependencies*,

- and Learnability*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Lightfoot, D. (1999) *The Development of Language*, Blackwell, Oxford, UK.
- Manzini, R. and K. Wexler (1987) "Parameters, Binding Theory, and Learnability," *Linguistic Inquiry* 18, 413-444.
- Pesetsky, D. and E. Torrego (2001) "T-to-C Movement: Causes and Consequences," in M. Kenstowicz, ed., *Ken Hale: A Life in Language*, MIT Press, Cambridge, Massachusetts.
- Reinhart, T. and E. J. Reuland (1991) "Anaphors and Logophors: An Argument Structure Perspective," in J. Koster and E. J. Reuland, eds., *Long Distance Anaphora*, Cambridge University Press, Cambridge.
- Reinhart, T. and E. J. Reuland (1993) "Reflexivity," *Linguistic Inquiry* 24, 657-720.
- Reuland, E. J. (2001) "Primitives of Binding," *Linguistic Inquiry* 32, 439-492.
- Safir, K. (1987) "Comments on Wexler and Manzini," in T. Roeper and E. Williams, eds., *Parameter Setting*, Reidel, Dordrecht.
- Stepanov, A. (2001) "Late Adjunction and Minimalist Phrase Structure," *Syntax* 4, 94-125.
- Suzuki, N. (1997) *Studies in Generative Syntax*, Eichosha, Tokyo.
- Watanabe, A. (1992) "Subjacency and S-Structure Movement of *Wh*-in-Situ," *Journal of East Asian Linguistics* 1, 255-291.
- Wexler, K. (1993) "The Subset Principle Is an Intensional Principle," in E. Reuland and W. Abraham, eds., *Knowledge and Language Volume 1: From Orwell's Problem to Plato's Problem*, Kluwer Academic Publishers, Dordrecht.
- Wexler, K. and P. Culicover (1980) *Formal Principles of Language Acquisition*, MIT Press, Cambridge, Massachusetts.
- Wexler, K. and R. Manzini (1987) "Parameters and Learnability in Binding Theory," in T. Roeper and E. Williams, eds., *Parameter Setting*, Reidel, Dordrecht.